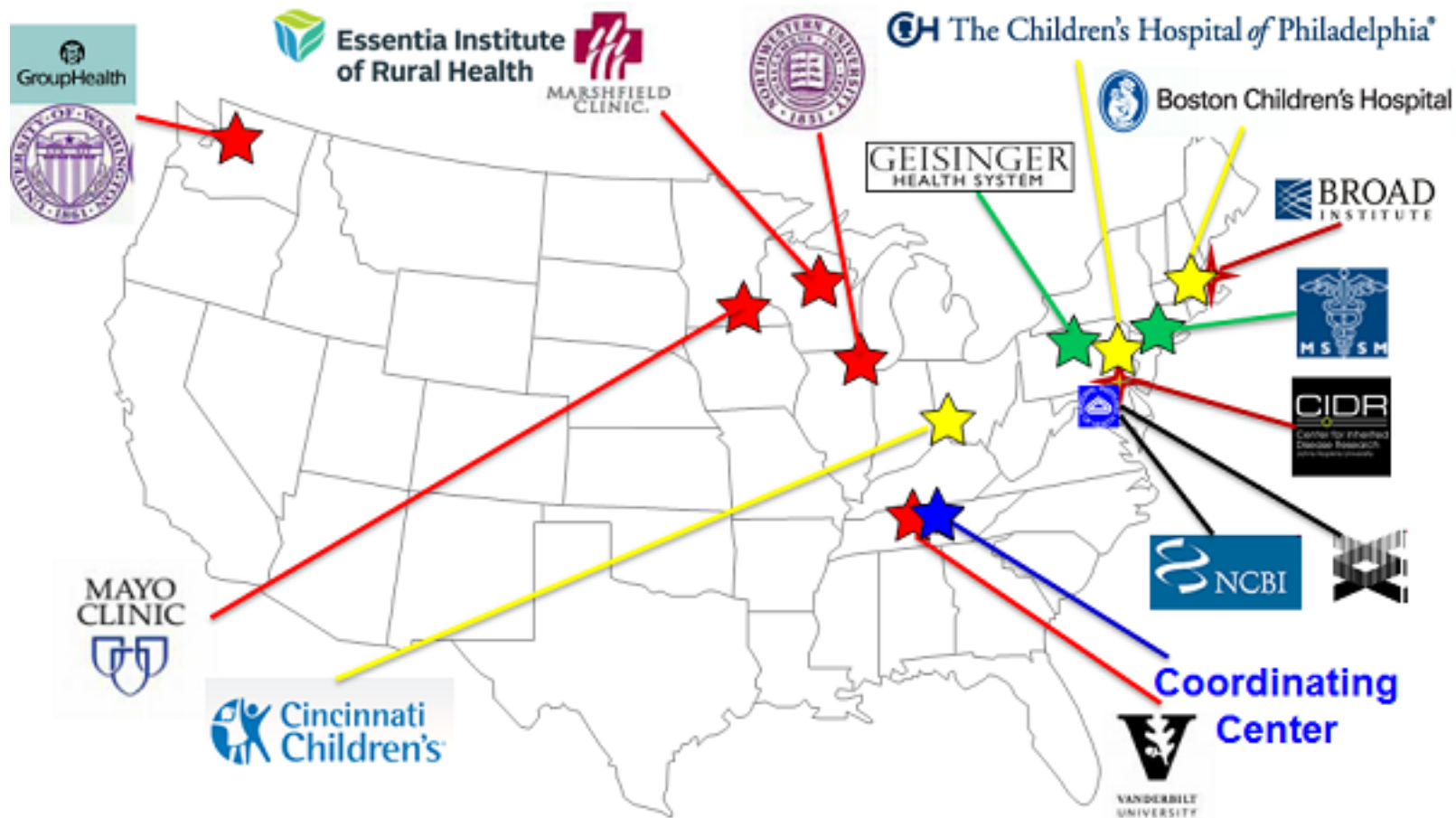


Data Sharing: Lessons from the eMERGE network

Rex L. Chisholm, PhD
Feinberg School of Medicine
Northwestern University
Chicago, IL USA



NHGI STRATEGIC PLAN - 2011

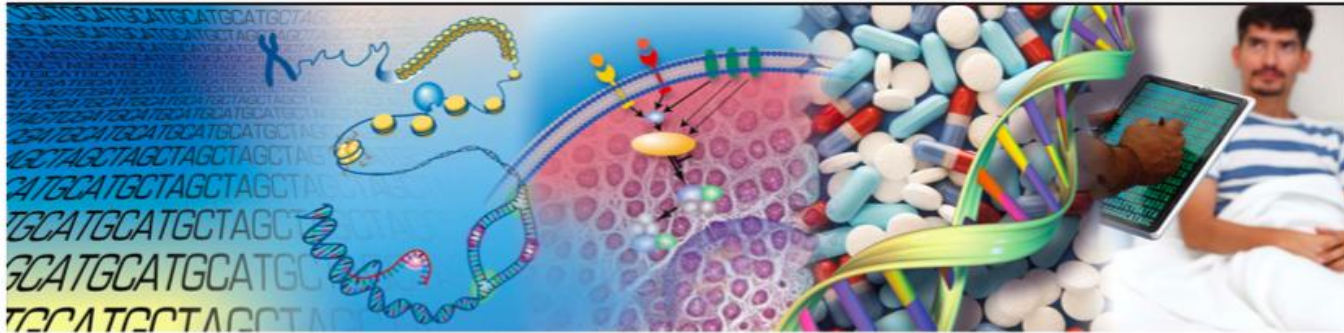
Understanding
the structure of
genomes

Understanding
the biology of
genomes

Understanding
the biology of
disease

Advancing
the science of
medicine

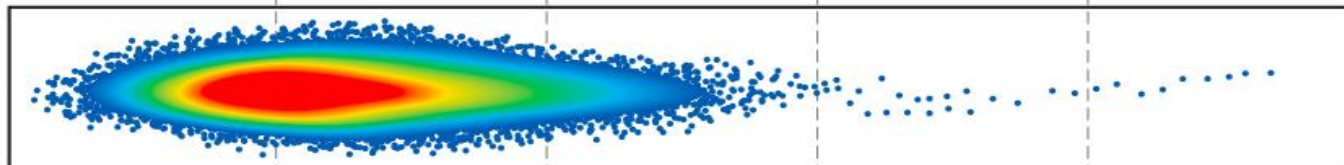
Improving the
effectiveness of
healthcare



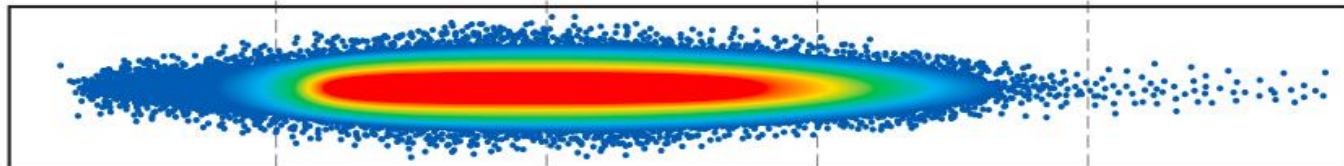
1990–2003
Human Genome Project



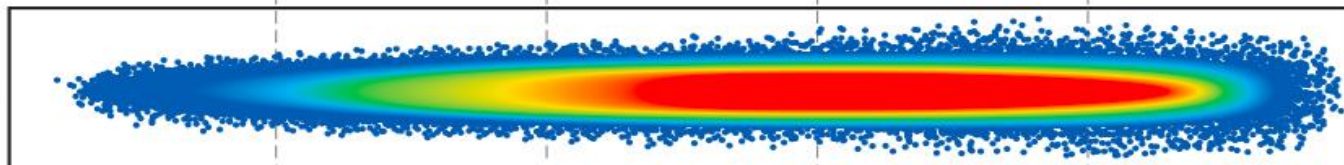
2004–2010



2011–2020



Beyond 2020



eMERGE Goals

- Phenotyping
- Genomics Genotype-Phenotype associations
- Clinical Use:
 - Defining actionability/clinical utility/validity
 - Integration into EHR/Visualization/Clinical Decision support
- Physician and Patient attitudes/Education
- Consent/Regulatory
- Privacy/Security/CLIA/CAP

eMERGE Sample Size

	eMERGE I		eMERGE II		eMERGE I & II
	Participants	Genotyped	Participants (Enrolled/ Targeted)	Genotyped	Genotyped
GHC/UW	2,820	2,789	3,561	786	3,575
Marshfield	20,000	4,210	20,000	777	4,987
Mayo	3,769	3,755	19,000	3,185	6,940
NU	10,500	1,907	10,500	3,055	4,962
VU	70,000	6,055	140,000	27,173	33,228
Geisinger	N/A	N/A	19,650	4,191	4,191
Mt. Sinai	N/A	N/A	21,000	16,000	16,000
CCMC/CHB	N/A	N/A	40,051	5,586	5,586
CHOP	N/A	N/A	40,000	8,000	8,000
	107,089	18,716	313,762	68,753	87,469

	GHC/UW	Marshfield	Mayo	Northwestern	Vanderbilt
Primary					
Dementia	X	X			X
Cataract		X			X
PAD		X	X	X	X
Type 2 Diabetes		X	X	X	X
QRS Duration		X	X	X	X
Secondary					
WBC	X	X	X	X	X
Diabetic Retinopathy	X	X			X
RBC	X	X	X	X	
Lipids		X	X	X	
Height		X	X	X	
PheWAS					X
HDL	X	X	X		
Network					
Hypothyroidism	X	X	X	X	X
Resistant HTN	X	X	X	X	X

eMERGE II working groups

- Phenotyping
- Genomics
- EHR integration (both technical and content)
- Consent, education, regulation and consultation
- Actionable variants

Work Groups to produce “charters”

Data Use Agreement

For use by and among Members of the Electronic Medical Records and Genomics Research Network (eMERGE)

TERMS AND DEFINITIONS: The Electronic Medical Records and Genomics (eMERGE) Network (<https://www.mc.vanderbilt.edu/victr/dcc/projects/acc/index.php/About>) is a National Institutes of Health (NIH)-organized and -funded consortium of ~~selected~~ research institutions ("eMERGE Network"). The primary goal of ~~the~~ eMERGE Network is to develop, disseminate, and apply approaches to research that combine DNA biorepositories with electronic medical record (EMR) systems for large-scale, high-throughput genetic research. Member institutions participating in the consortium study the relationship between genetic variations and clinically relevant human traits, using the technique of genome-wide association (GWAS) analysis. Such studies involve testing hundreds of thousands of genetic variants called single nucleotide polymorphisms throughout the genome in people with and without a condition of interest. A fundamental question that eMERGE seeks to answer is whether electronic medical record (EMR) systems can serve as resources for such complex genomic analysis of disease susceptibility and therapeutic outcomes, across diverse patient populations. In addition, the consortium includes a focus on social and ethical issues such as privacy, confidentiality, and interactions with the broader community. Detailed information on ~~the~~ eMERGE network can be found at the eMERGE website (www.eemr.org).

eMERGE Network members ("eMERGE Network Members" or "Members") include: sites funded through the primary eMERGE Network grant mechanism, the Network's coordinating center, the National Center for Biotechnology Information (NCBI), NIH; and the National Human Genome Research Institute (NHGRI), NIH, and other sites with whom the Members have agreed to partner with in accordance with the eMERGE Network Criteria for Participation and which have signed this Data Use Agreement.

Researchers with a wide range of expertise in genomics, statistics, ethics, informatics, and clinical medicine employed by an eMERGE Network Member participate in the eMERGE network, including: Principal Investigators of the eMERGE clinical sites, the Coordinating Center, the Genotyping Facilities, and Program Officials from the National Center for Biotechnology Information and the National Human Genome Research Institute.

Data Sharing Guiding Principles: All data sharing will adhere to 1) the terms of consent agreed to by research participants; 2) applicable laws and regulations; and 3) the principle that individual sites within the network have final authority regarding whether their site's data will be used or shared, on a per-project basis. These principles are intended to maximize sharing of GWAS data generated by the eMERGE Members among and between other Members as well as with the wider scientific community, and to do this without compromising data security or the confidentiality of information about individuals whose data and/or samples are used for research.

Data Sharing Responsibilities: Principal Investigators of each eMERGE clinical site may designate data to accomplish activities defined in research studies sanctioned by the eMERGE Network according to eMERGE Network policies (eMERGE Data) to be shared as follows: (1) distribution through dbGaP; (2) distribution within the eMERGE Network; and/or (3) distribution to the eMERGE Coordinating Center. ~~The~~ eMERGE Data to be shared within eMERGE will be provided only to eMERGE Network Members that have signed this Agreement. All eMERGE Network Members and the eMERGE Coordinating Center may aggregate eMERGE Data from all Member sites and, with documented approval prior to each submission from the contributing site(s), submit said eMERGE Data to dbGaP and/or other databases administered by the National Institutes of Health. Each eMERGE Network Member may share its own data with external collaborators without approval of the other Members. If eMERGE data received from any Member is shared externally by another Member, prior approval from the Member providing ~~the~~ eMERGE Data must be obtained and documented. Members sharing eMERGE Data externally must also ensure that each external eMERGE Data recipient agrees to the same restrictions and conditions applicable to Members and Member Representatives regarding the use and disclosure of the eMERGE Data as outlined in this Agreement or as may be required by law. ~~The~~ eMERGE Data may be shared by Members with their partner organizations (e.g., organizations that provide direct services such as genotyping on Member-provided samples, site-initiated subcontracted services and/or other services required for eMERGE sanctioned studies), provided that said partner organizations are held to an equivalent standard of confidentiality

that applies to Network Members

(https://www.mc.vanderbilt.edu/victr/dcc/projects/acc/index.php/Member_Resources).

Statement of Confidentiality: By signing this Agreement, the authorized official representing an eMERGE Network Member, certifies that s/he and the Principal Investigators, fellows, students, and research staff (collectively, "Network Member Representatives") working on eMERGE-related projects are aware of the confidential nature of data on research participants maintained by the Member and of the necessity for maintaining that confidentiality.

~~The~~ eMERGE Network Member agrees not to attempt to personally identify any eMERGE participant based on eMERGE Data and agree not to attempt to contact any eMERGE participant of a site other than their own. The Member agrees not to transfer or disclose any confidential data or any information about individual eMERGE participants, except as permitted by this Agreement or as required by law, either during or after the conclusion of the affiliation with eMERGE. The Member agrees to provide adequate security for the eMERGE Data, including but not limited to safeguards intended to prevent unauthorized use or disclosure of such information. In addition, each Member agrees to report in writing to the other Members any use or disclosure of any portion of the data of which it becomes aware that is not permitted by this Agreement including disclosures that are required by law.

The eMERGE Network Member agrees to ensure that its Network Member Representatives do not use, disclose or transfer any eMERGE Data to anyone who is not an eMERGE Network Member except as permitted by this Agreement or as required by law. Further, the Member agrees to return all eMERGE Data to the eMERGE Coordinating Center or delete/destroy all electronic eMERGE Data upon termination of its affiliation with the eMERGE Network and to notify ~~the~~ eMERGE Coordinating Center when it has done so.

Limitations of Data Use: The eMERGE Network Member agrees to ensure that Network Member Representatives will only use eMERGE Data in a manner that is consistent with any limitations that have been specified for individual studies by the disclosing Member and agreed to by the Steering Committee and shall ensure compliance with all applicable state and federal laws and regulations governing the use of such data including the Health Insurance Portability and Accountability Act of 1996 (HIPAA), if applicable, including any and all future amendments.

The eMERGE Network Member agrees to comply with all established policies of eMERGE governing the acquisition, analysis, reporting, publication, use and distribution of eMERGE Data.

This Agreement supersedes and replaces all prior agreements made between eMERGE Network Members regarding the sharing and confidentiality of eMERGE Data.

Agreed to by:

~~eMERGE~~ Institution Authorized Official name and title (print): _____

Signature: _____ Date: _____

Read & Understood by:

Network Member Representative name & title (print): _____

Signature: _____ Date: _____

Member Representative's Institution:

Data Sharing in eMERGE

Guiding Principles:

- Adhere to terms of consent
- Adhere to applicable law and regulations
- Each site has final authority regarding their data

Data Sharing in eMERGE

Data Sharing Responsibilities:

How data may be shared:

- To dbGaP
- To and by other Network Members
- To and by the CC
- To external collaborators

Data Sharing in eMERGE

Privacy and Confidentiality

- Prevents re-identification
- Provides for data security
- Reporting requirements for violations
- Provides for data protections in use
- Requirements for data return/destruction if Member affiliation is terminated

Limitations of Use

Authorizations by institutional official and PI

The logo for the PhenX Toolkit features three stylized human figures in grey on the left, followed by a blue DNA double helix. To the right of these icons, the word "PhenX" is written in a large, bold, black sans-serif font, and the word "Toolkit" is written in a smaller, gold-colored sans-serif font.

PhenX Toolkit

- Provides standard measures related to
 - complex diseases
 - phenotypic traits
 - environmental exposures
- Facilitates combining data from a variety of studies
- Makes it easy for investigators to expand a study design beyond the primary research focus.

Browse Domains

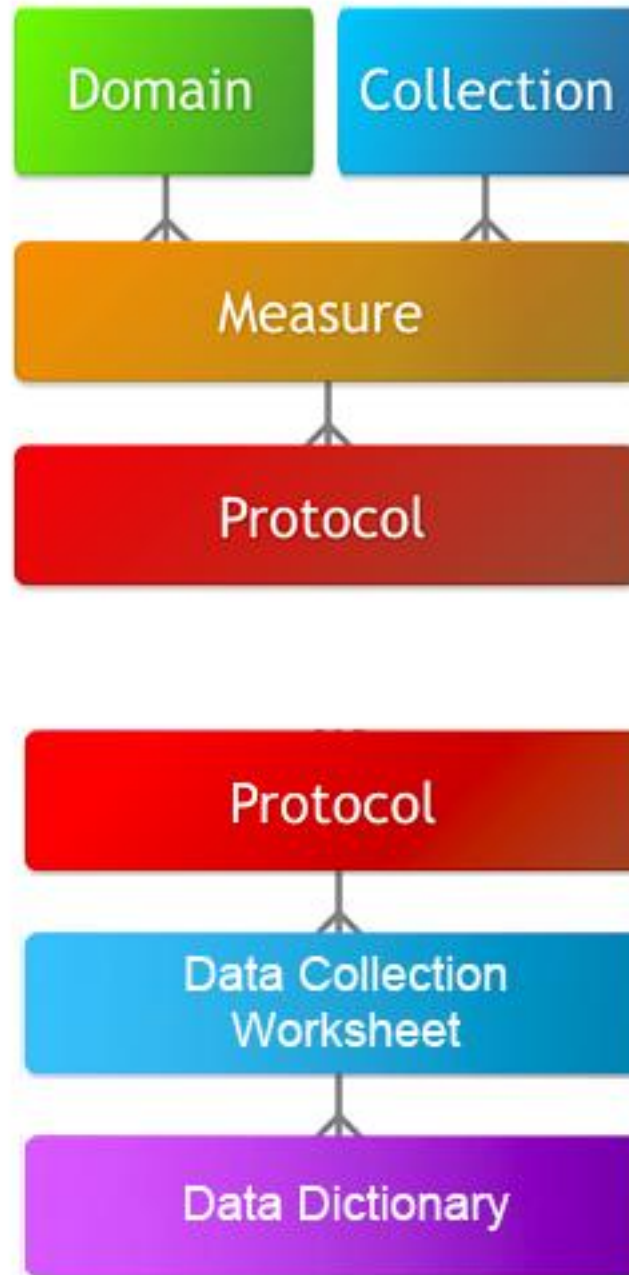
Show Tree

There are a total of 339 measures in the PhenX Toolkit. Browse through Domains to view Measures and Protocols. You may also [Browse Measures](#) » and [Browse Collections](#) »

- | | |
|---|---|
| Add to My Toolkit #030000 Alcohol, Tobacco and Other Substances (14) » | Add to My Toolkit #050000 Nutrition and Dietary Supplements (12) » |
|  Substance Abuse and Addiction » | Add to My Toolkit #110000 Ocular (15) » |
| Add to My Toolkit #020000 Anthropometrics (16) » | Add to My Toolkit #080000 Oral Health (15) » |
| Add to My Toolkit #070000 Cancer (12) » | Add to My Toolkit #150000 Physical Activity and Physical Fitness (14) » |
| Add to My Toolkit #000000 Cardiovascular (14) » | Add to My Toolkit #120000 Psychiatric (14) » |
| Add to My Toolkit #010000 Demographics (15) » | Add to My Toolkit #180000 Psychosocial (15) » |
| Add to My Toolkit #140000 Diabetes (15) » | Add to My Toolkit #100000 Reproductive Health (15) » |
| Add to My Toolkit #060000 Environmental Exposures (14) » | Add to My Toolkit #090000 Respiratory (14) » |
| Add to My Toolkit #190000 Gastrointestinal (12) » | Add to My Toolkit #170000 Skin, Bone, Muscle and Joint (10) » |
| Add to My Toolkit #160000 Infectious Diseases and Immunity (15) » | Add to My Toolkit #210000 Social Environments (15) » |
| Add to My Toolkit #130000 Neurology (14) » | Add to My Toolkit #200000 Speech and Hearing (15) » |

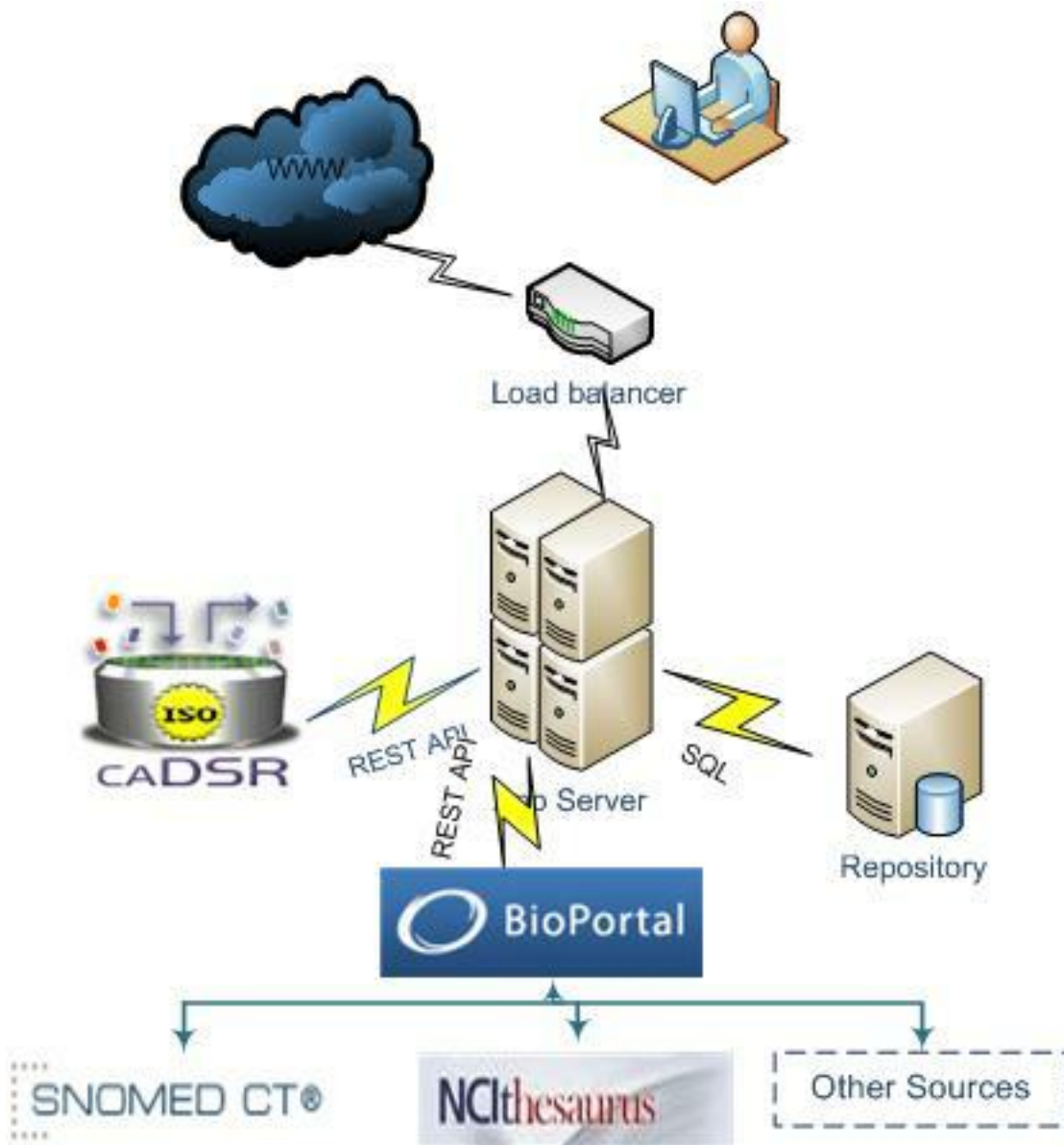
The number of Measures in each Domain is shown in parentheses.

The numbers in the #number format indicate unique codes for each domain, measure and protocol.



eleMAP

harmonizes
phenotype
data
dictionaries to
existing
metadata and
terminology
standards





**PUBLIC
POPULATION
PROJECT IN
GENOMICS
AND SOCIETY**

[USERNAME]
[PASSWORD] OK
[Request an account](#)
[Request new password](#)

Biobank
Lexicon

- About P3G
- Resources
- Programmes
- Membership
- Events
- News
- Contact Us



LOOKING FOR OPPORTUNITIES TO LEARN AND NETWORK WITH EXPERTS?

👉 Browse through our list of International Events

 <p>TOOLKIT SEARCH BIOBANKING TOOLS BY NAME, CATEGORY OR TYPE</p>	 <p>LIFESPAN BROWSE THROUGH THE DIFFERENT PHASES AND STEPS OF THE BIOBANKING LIFECYCLE</p>	 <p>HUB DISCUSS, EXCHANGE AND COLLABORATE WITH EXPERTS AROUND THE WORLD</p>	 <p>TRAINING ACCESS TUTORIALS AND INFORMATION SESSIONS</p>	 <p>CATALOGUES SEARCH FOR INFORMATION ABOUT POPULATION-BASED BIOBANKS</p>	 <p>BRIF REQUEST/SEARCH FOR A BIORESOURCES RESEARCH IMPACT FACTOR</p>
---	--	---	--	---	---

First time on our new site? Visit the [Resources](#) section to learn more about these modules

EVENTS

NEWS



Is rigorous retrospective harmonization possible? Application of the DataSHaPER approach across 53 large studies

Isabel Fortier,^{1,2*} Dany Doiron,² Julian Little,^{1,3} Vincent Ferretti,⁴ François L'Heureux,² Ronald P Stolck,⁵ Bartha M Knoppers,^{1,6} Thomas J Hudson,^{4,7,8} and Paul R Burton^{2,9,10} on behalf of the International Harmonization Initiative[†]

¹Research Institute – McGill University Health Centre, Montreal, Quebec, Canada, ²Public Population Project in Genomics (P³G), Montreal, QC, Canada, ³Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, ON, Canada, ⁴Ontario Institute for Cancer Research, MaRS Centre, Toronto, ON, Canada, ⁵Department of Epidemiology, University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands, ⁶Department of Human Genetics, Centre of Genomics and Policy, Faculty of Medicine, McGill University, Montreal, QC, Canada, ⁷Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada, ⁸Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada and ⁹Department of Health Sciences, University of Leicester, Leicester, UK and ¹⁰Department of Genetics, University of Leicester, Leicester, UK

*Corresponding author. Research Institute – McGill University Health Centre, Allen Memorial Building, 1025 Pine Avenue West, room P2.028, Montreal, Quebec, Canada, H3A 1A1. E-mail: ifortier@p3g.org

[†]The members of the International Harmonization Initiative are provided in Appendix.

Summary

- eMERGE benefited from effective and easy data sharing
 - Synergies increase the scientific value
 - More perspectives are always better than just one
- Sharing requires agreed upon principles and shared values
- Sharing requires data standards and data harmonization
- Try it—you'll like it!!

Breakout Sessions

Suggested chairs (H3A PIs) for 3 break out groups:
Akin Abayomi (biorepository), Dissou Affolabi
(research grant) Enock Matouvo (research grant)

Staff Note takers:

Louise Wideroff, Mark Guyer, Audrey Duncanson

Questions for sharing breakout groups

1. Are there specific needs in Africa for data and protocol sharing that have not already been discussed?
2. What, in your opinion, would be a fair timeline for data release?
3. What protocols would be useful for H3Africa grantees to share? How rapidly should they share them?
4. What can be done to make data sharing and deposition easier?